

ABSTRACT

[0104] A nucleic acid fragment of interest is incorporated into a hybrid artificial gene and expressed in one or more reading frames to produce one or more hybrid polypeptides. The polypeptides are examined with respect to one or more physical parameters, such as mass or amino acid composition. The observed parameter values are used to search a data set of predicted parameter values generated by hypothetical translation of a larger reference nucleic acid sequence so as to determine whether or not the fragment is contained within the reference sequence, and, if it is contained therein, to determine its sequence and/or coding capacity.